



van der Linden, D., Zamansky, A., & Hadar, I. (2017). A framework for improving the verifiability of visual notation design grounded in the physics of notations. In *Requirements Engineering Conference (RE), 2017 IEEE 25th International* (pp. 41-50). (Requirements Engineering, IEEE International Conference). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/RE.2017.37>

Peer reviewed version

License (if available):
Other

Link to published version (if available):
[10.1109/RE.2017.37](https://doi.org/10.1109/RE.2017.37)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via IEEE at <https://doi.org/10.1109/RE.2017.37> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

A framework for improving the verifiability of visual notation design grounded in the Physics of Notations

Dirk van der Linden, Anna Zamansky, and Irit Hadar

Department of Information Systems

University of Haifa

Haifa, Israel

{djtlingen,annazam,hadari}@is.haifa.ac.il

Abstract— This paper proposes a systematic framework for applying the Physics of Notations (PoN), a theory for the design of cognitively effective visual notations. The PoN consists of nine principles, but not all principles lend themselves equally to a clear and unambiguous operationalization. As a result, many visual notations designed according to the PoN apply it in different ways. The proposed framework guides what information is required of a reported PoN application to ensure that the application of each principle is verifiable. The framework utilizes an evidence-driven design rationale model to structure information needed to assess principles requiring user involvement or cognitive theories. This approach aims to reduce ambiguity in some of the principles by making design choices explicit, and highlighting the level of evidence presented to support it. We demonstrate the proposed framework in a showcase of a recently published visual notation which has been designed with the PoN in mind.

Index Terms— conceptual modeling, visual notations, cognitive effectiveness, design rationale, physics of notations

I. INTRODUCTION

Visual notation design is an important part of the design of any modeling language, as the visual notation is the main interface of its users to any produced model. Given the use of conceptual modeling to foster communication among stakeholders in development processes, it is of great importance to ensure that a notation is designed in such a way that it can be quickly and correctly understood by all those involved. In other words: that it is cognitively effective. A number of approaches exist to support designers in doing so, one of which in particular has received more attention than others over the last years: The Physics of Notations (PoN) [4], a theory aimed to ensure that visual notations are designed to be cognitively effective. At least twenty-three *new* notations designed with it in mind have been published [9]. The PoN consists of nine principles, all of which have to be addressed in the design of a notation to ensure its cognitive effectiveness. Table 1 presents these principles and a short description explaining each of them. It is important to note, however, that each of the principles encompasses a far deeper wealth of theoretical and practical considerations than their one-line summary would inspire from a naïve reading.

Perhaps because of this depth and complexity, the PoN has received some critique in the literature, mainly focused on the difficulty of systematically applying it. Da Silva Teixeira et al. [5] proposed an additional systematic process to help designers chose when to apply which principle, grouping together those principles whose implementation goes hand in hand. However,

TABLE 1. OVERVIEW OF THE PoN'S PRINCIPLES AND THE ABBREVIATIONS USED FOR THEM AFTERWARDS.

Principle	Abbrev.	Description (from [2])
Semiotic clarity	SemCla	there should be a 1:1 correspondence between semantic constructs and graphical symbols
Perceptual discriminability	PerDis	symbols should be clearly distinguishable from one another
Semantic transparency	SemTra	use symbols whose appearance suggests their meaning
Complexity management	ComMan	include explicit mechanisms for dealing with complexity
Cognitive integration	CogInt	include explicit mechanisms to support integration of information from different diagrams
Visual expressiveness	VisExp	use the full range and capacities of visual variables
Dual coding	DuaCod	use text to complement graphics
Graphic economy	GraEco	keep the number of different graphical symbols cognitively manageable
Cognitive fit	CogFit	use different visual dialects for different tasks and/or audiences

this process does not provide more guidance on how to actually apply the principles themselves, leaving it up to the interpretation of the designer. Störrle and Fish [6] started an effort to provide clearly formalized definitions of the principles, yet have only provided definitions for semiotic clarity and perceptual discriminability so far – and even there admitted to making arbitrary choices for some thresholds. More recent work proposed theory-grounded color palettes for perceptual discriminability [7]. However, those principles more strongly linked to human factors related to cultural or individual aspects (e.g., semantic transparency requiring understanding of connotative meaning) were passed over as of yet – an issue shared by a recently proposed editor designed to enable PoN-driven development of visual notations [8]. Indeed, the PoN principles were strongly criticized in general as being “neither precise nor comprehensive enough to be applied in an objective way” [6].

This seeming different level of how well each principle can be operationalized led us to investigate the degree to which the principles lend themselves to such formalization efforts [9]. We had to conclude that it is infeasible, if not downright impossible to capture several of the principles in finite formalizations. An important reason for these difficulties is the strong need for user involvement in verifying whether principles hold, even though user involvement in PoN applications is often lacking

[10]. Involving users requires extensive empirical work, likely leading to different outcomes in different contexts-of-use, thereby reducing the generalizability of any instructions that could be given for these principles. For example, while an effective color scheme could be created to satisfy a part of visual expressiveness [7], depending on the actual people using it, the instantiations of that scheme would be different, such as between e.g., Germanic and Confucian cultures [11]. A formalization, therefore, would either necessarily have to admit to only being valid for a very particular context-of-use (including its target audience and task), or present specific instantiation options for every feasibly expected different context-of-use, a rather unlikely achievement to be made.

What, then, is the way forward in guiding designers to apply the PoN in a systematic way? We see this as a call for more systematic reporting on the side of the designer of the visual notation. Indeed, in one of the first major applications of the PoN performed by Moody himself in 2010 [2], he stressed the importance of providing explicit design rationale. Explication of design rationale, and its grounding in verifiable sources is one of the main things needed to understand – and assess – the choices made in a visual notation’s design. Doing so would also strengthen the need to understand the spirit of each principle and designing with that in mind, as opposed to paying mere lip service to it. Nearly seven years later, it seems that this lesson has not yet resounded as loudly as hoped.

The objective of this paper is to provide a more systematic way to apply the PoN, guiding designers to more effectively implement the principles, and how to report on it. Avoiding the pitfall of attempting to formalize every last minutia of each principle, we propose a model for design rationale specific to this purpose, by which the credibility of a notation’s claim that any or all principles are satisfied can be validated – and where necessary, strengthened – by others. To do so we will first assess what an exemplary application of the PoN presents in their reasoning for each principle’s satisfaction, and derive a structured framework for what ought to be presented, and how justifications for design choices should be structured according to an evidence-focused model for design rationale.

The rest of this paper is structured as follows. In Section 2 we analyze an exemplary application of the PoN, taking the i^* paper by Moody et al. [2] as our case study to assess what is reported in a thorough example of a PoN application. Section 3 goes into more detail on what is needed to capture relevant detail of design rationale for our purposes. This is followed in Section 4 by the presentation of the framework for systematically applying the PoN principles. We apply it to a recently published visual notation in Section 5, followed by a discussion on limitations and considerations on the framework in Section 6, and a concluding outlook for further work in this direction in Section 7.

II. VERIFIABILITY OF PoN APPLICATIONS: RECONSTRUCTING REPORTING REQUIREMENTS

A. General approach

Our first step is to construct a set of requirements for ensuring the verifiability of design grounded in the PoN. In other

words: what does a designer actually have to report in order to make it possible for others to verify that the proposed design is cognitively effective according to the PoN? Given the body of work that has applied the PoN while designing a new visual notation (cf.[12]), it should be possible to reconstruct a set of requirements by investigating detailed applications of the PoN. In particular, we start by investigating the application of the PoN to the goal modeling notation i^* [2]. We chose this particular application as a starting point because being authored by the PoN’s creator, it is closest in spirit to the intended operationalization of the principles. In addition, it includes additional reflections on what is important in applying the theory. First we will briefly discuss what the authors actually state in regards to each PoN principle, and then summarize what (kinds of) information is needed to support these statements.

B. What is reported

Semiotic clarity: starts with a presentation of (i) all semantic constructs of i^* , and (ii) all its visual constructs. Based on (i) and (ii), calculations are made to determine the degree of symbol redundancy, overload, excess and deficit.

Perceptual discriminability: starts with a presentation of all shapes used by the visual constructs in i^* (that is, the different instantiations of the visual variable ‘shape’). Based on this, several aspects (shape similarity, shape inconsistency and discriminability of relationships) are analyzed. *Shape similarity:* the visual similarity between some shapes (goal/belief and agent/roles) are pointed out. It is noted that this similarity can lead to confusion, based on experimental studies on ER diagrams which showed that novices confuse between similar shapes of triangles and diamonds. *Shape inconsistency:* it is noted that POSITION, a subtype of ACTOR is represented by a shape from a different family (a rounded-up diamond compared to a circle). A possible design rationale for this choice is mentioned, and the lack of its documentation is highlighted. *Discriminability of relationships:* the authors chose to analyze relationships separately from entities. Textual differentiation is noted to be used extensively in i^* relationships, and deemed cognitively ineffective, citing a previous article which found similar issues with the UML notation.

Semantic transparency: starts with a presentation of the scale of semantic transparency as defined by the PoN (see Fig. 1). The visual constructs of i^* are reviewed against this scale, concluding that the majority is semantically opaque, being abstract geometrical shapes (with the exception of the cloud symbol used for BELIEFS.)

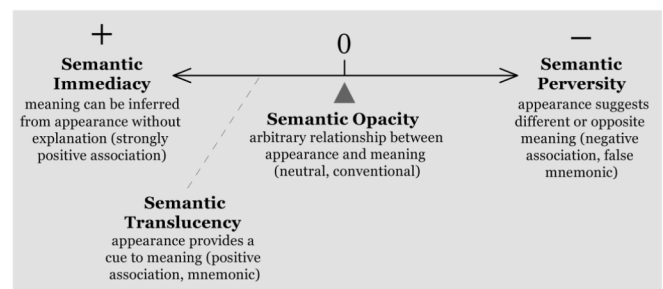


Fig. 1. Transparency scale, from [4].

Complexity management: starts with noting the number of levels of abstraction allowed by i^* constructs (two). It is then noted that i^* lacks a way of breaking diagrams into modules or chunks (using recursive decomposition), ensuring monolithic diagrams regardless of complexity. This implies a low level of complexity management mechanisms.

Cognitive integration: is analyzed by assessing the existence of supporting mechanisms for multi-diagram representations, namely (i) cognitive integration and (ii) perceptual integration. Cognitive integration is not deemed a major problem because i^* only has two types of diagrams (strategic dependency and strategic rationale). Following this, naming conventions for the diagram types are analyzed, noting that they are too similar. Perceptual integration is noted to be affected by the use of different symbols to represent the same concept (ACTORS) on the different diagram types. Some strengthening of perceptual integration is provided by Strategic Dependency (SD) diagrams in i^* which can be viewed as long-shot diagrams. However, the authors state that SD diagrams are not effective in this sense, citing literature to support their view.

Visual expressiveness: starts with listing all visual variables used in i^* : shape, brightness and orientation, determining a visual expressiveness score of 3. The visual expressiveness is noted to be strengthened by the use of curved shapes, with the authors noting works claiming that these are more perceptually efficient and aesthetically pleasing. The authors then consider the effectiveness of using the visual variable of color. They evaluate it as poor in i^* based on the following considerations: (i) color does not contribute to perceptual discriminability, i.e., differentiating between symbols (different shapes have same color) and relationships (arcs use text instead of color), (ii) color is used on both text and background. The authors claim that according to literature this reduces legibility and aesthetics.

Dual coding: starts by checking whether text is used to complement visual constructs (both entity or relationships). It is noted that i^* does not combine text and visual constructs.

Graphic economy: the number of different visual constructs are counted and compared to the established cognitive upper limit of 7 ± 2 symbols. i^* is noted to have a graphic complexity of 16 symbols, highly exceeding the recommendation.

Cognitive fit: the number of visual dialects is counted to assess whether enough dialects are available to support optimal representation for different users and representational media. i^* is noted to have no such dialects.

C. What information is used to report

A critical piece of information for several of the principles is a complete representation of the semantic and visual constructs. For example, for semiotic clarity these constructs are presented first, after which symbol redundancy, overload, excess and deficit calculated. Verifying the size of the visual vocabulary for graphic complexity also depends on having this set of visual constructs available to the reader. The analysis of each visual construct's location on the semantic transparency scale depends on the symbols being available to the reader as well. Thus, for an application of the PoN to be verifiable at all it is necessary to explicitly present the semantic and visual con-

structs, and clear labeling of both so readers can link them. Visual constructs can be presented simply as listing of all used symbols (entities and relationships alike), while semantic constructs can take the form of e.g., meta-models, class hierarchies, or textual definitions. For perceptual discriminability the set of used shapes is also needed. Although they follow from the set of visual constructs, for simplicity's sake they can be reported on explicitly.

Some principles pose requirements on the diagram level, where it is not enough to only look at atomic constructs. For example, for cognitive integration we need to know whether we can create relationships between entities in distinct diagrams. To assess complexity management, we need to not only know whether there are semantic and visual constructs for abstraction (e.g., black boxes, collapsible elements), but also how these elements behave: knowledge of the compositional rules for forming valid expressions are necessary. In some cases, this information is already encoded in the presentation of the semantic constructs (e.g., in a given meta model or specification); in other cases it still needs to be explicitly reported on.

The PoN itself is grounded in relevant theories from a number of fields (e.g., semiotics, psychophysics, cognitive science), but knowledge of those theories is not always needed. For example, to apply semiotic clarity, simply counting the occurrences of symbol redundancy, overload, excess and deficit can be done without having an in-depth understanding of *why* these ought to be avoided. This is not the case for all principles, some of which are strongly linked to the application of a specific cognitive theory. For example, the PoN's normative instructions in terms of dual coding are a near-straightforward statement to apply dual-coding theory to the visual notation at hand. Given the PoN's ambiguous operationalization which does not make clear how much thought should be put into this dual coding, i.e., "use text to supplement visual information", it becomes necessary for the designers to familiarize themselves with the relevant theory and decide exactly what to do. This leads to a certain variability in how the PoN is wielded, making it very important to understand how the designer wielded these theories. To make it verifiable what was done, the rationale for what parts of the theory were applied, and in how much detail, is needed. The principles that are closely linked to a cognitive theory, without giving explicit instructions, cognitive integration, cognitive fit, and dual coding all require such details – a finding shared by our earlier work on a minimal operationalization of PoN principles [9].

Design rationale is not only important to clarify how and to what degree particular theories were applied, but also to directly ground decisions taken for individual visual properties. For example, when determining perceptual discriminability, the similarity of shapes has to be assessed, which is not a matter of clear-cut yes or no questions. Say we have two constructs, GOAL and SOFTGOAL, which are represented as \otimes and \times respectively. Instead of simply stating that they are dissimilar (enough), we have to provide a rationale *why* we believe they are dissimilar. Here, for example, a naïve, but commonly used kind of claim would be that the difference in font and slight incline makes them distinct enough. A more substantial

grounding would involve asking users to chose the two most distinct “x” characters from a list and using the most provided answer. For several principles such design rationale is important. For example, in the i* assessment, Moody proposes to increase visual expressiveness via color usage, suggesting using yellow for tasks by association with sticky notes, resource as green by association with trees, soft goals as pink by association with “softness or fluffiness.” All these design choices are given with a particular *design rationale*. For semantic transparency design rationale is of particular importance, as the claim for each symbol suggesting its meaning is dependent on many personal and cultural factors, making it important to verify that such design holds in a given context-of-use.

D. Summary: information required for verifiability of PoN applications

The analysis discussed above can be summarized into an overview of what data – semantic constructs, visual constructs, rules for linking constructs, and design rationale – are needed to verify each principle. An overview is given in Table 2, where each × represents some information that should be reported for a PoN application’s design claims to be verifiable.

TABLE 2. MINIMAL INFORMATION NEEDED TO VERIFY IF A PoN PRINCIPLE HOLDS, DERIVED FROM SEC. II.B AND II.C. CONSTRUCTS MARKED WITH AN ASTERISK (*) REQUIRE EXPLICATION OF THE COMPOSITIONAL RULES FOR CONSTRUCTS OF THAT TYPE. DESIGN RATIONALE MARKED WITH A DAGGER (†) REQUIRE RATIONALE ON THE IMPLEMENTATION OF COGNITIVE THEORY.

Principle	Semantic constructs (S)	Visual constructs (V)	Design Rationale
SemCla	×	×	
PerDis	×	×	×
SemTra		×	×
ComMan		×	
CogInt	×	×	×†
VisExp		×	×
DuaCod	×		×†
GraEco		×	
CogFit		×	×†

Furthermore, as several principles require designers to adapt relevant cognitive theories to their requirements, in Table 3 we summarize which principles require familiarization with which additional theory. Besides such familiarization, these principles also require explicit design rationale for the way and degree in which these theories were applied.

TABLE 3. PRINCIPLES WHICH REQUIRE MORE IN-DEPTH FAMILIARIZATION WITH A COGNITIVE THEORY.

Principle	... requires familiarization with	i.e.
CogInt	Cognitive integration of diagrams	[13]
DuaCod	Dual-coding theory	[14, 15]
CogFit	Cognitive fit theory	[16]

III. REQUIREMENTS FOR CAPTURING DESIGN RATIONALE

The analysis summary in Table 2 shows the importance of design rationale for verifying PoN applications, as information about design rationale is needed for six out of nine principles. While semantic and visual constructs (and related linking rules) are typically straightforward to report on, design rationale requires more effort to be structured and presented systematical-

ly. Even in the exemplary application of the PoN we discussed, some presented design rationales are little more than personal opinion or grounded in gut feeling, making it clear that something is needed to stimulate more in-depth treatment to construct (and report on) rigorous design rationales.

Moody explicitly linked the concept of design rationale to the PoN concepts in the i* analysis (see Fig. 2). In particular, he postulated that design rationale should be specified at the lowest level of granularity on which the principles operate (visual properties), and most importantly, that such rationale should be based on “theory and/or empirical evidence rather than common sense, opinion or personal taste.” [2]

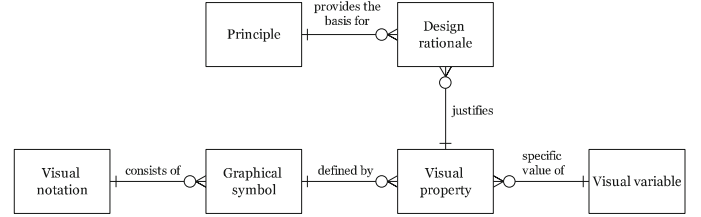


Fig. 2. Moody’s positioning of design rationale, from [2].

The model presented in Fig. 2, however, does not present enough detail or granularity to aid designers in structuring design rationale – nor does it stimulate designers to provide rationales based on evidence rather than opinion. We need a clearer structure to capture (or reverse engineer) the design rationale used to justify design choices made during the implementation of a PoN principle.

A. Design rationale: argumentation vs. evidence

The literature on design rationale encompasses an extensive set of options for doing so. However, we need to clarify the exact purpose for which we wish to capture design rationale in order to determine what information is most relevant [17]. Shipman and McCall [18] noted three major purposes for design rationale: argumentation, documentation and communication. As we are interested in ensuring that PoN-based design is verifiable by others, we are focused on the documentation purpose, which is the use of design rationale as an enabler for those outside the design process to understand how a design came to be [18]. For this purpose, the results of the reasoning process (i.e., the design choice) and the immediate explanation are most important, while other typical information such as alternative design choices are not as important.

The move towards an evidence-driven focus of design rationale rather than a traditional argumentation-driven focus, we should incorporate a way of distinguishing between the strength of different levels of evidence. Clinical research has developed many frameworks to capture the different degrees to which information represents evidence. A widely spread notion is that of Levels of Evidence (LoE). Sacket [19] introduced a model to distinguish between five levels of evidence, which has been built on by many others. In general, these earlier models make the distinction between evidence based on personal (or “expert”) opinion, correlational evidence, and causal evidence.

Fig. 5. Meta-model of our conceptual framework.

constructs, visual constructs, and their relevant **compositional rules**, and **rationale**. In the model **rationale** can be seen to serve a double role as the information that rationalizes a particular design choice, and information that is required to ensure the verifiability of that design choice. The structural form of a **rationale** is elaborated on below in Sec. III.C.

We make an explicit distinction between design choices and rationale so to be able to capture the scenarios where particular designs are made without rationale, or without that rationale being reported. Notably, in our approach PoN Principles do not provide the basis for design rationale as in Moody’s initial meta-model, because not all principles require design rationale in the same way. Furthermore, those principles that require explicit design rationale tend not to provide unambiguous, operationalized ways to achieve it, such as semantic transparency’s instruction to ensure symbols suggest their meaning.

To differentiate between how well grounded different **rationale** is, each rationale has to provide **evidence**, together with its **evidence level** as determined by Table 4. Typically, evidence falls into one of four categories, from weak to strong: ‘**expert**’ **evaluation**, **weak empirical**, **strong empirical**, and (well-established) **theory**. Examples of different kinds of evidence are given in Table 4. **Theory** and **PoN principle** are linked via a **is required to implement** relationship, as some principles (cognitive fit, dual coding, cognitive integration) explicitly require a cognitive theory to be implemented.

C. Capture profile


From the above meta-model we can now create a general profile for capturing the design rationale of design choices made during the design phase of a visual notation. When filled out, these profiles give a clearly structured overview of the reason and evidence for a particular design choice, which makes it more straight forward to verify and assess how likely they are to be valid instantiations of a principle.

TABLE 5. CAPTURE PROFILE FOR DESIGN CHOICE RATIONALE.

Title	Descriptive title
Principle	The principle which the design choice is made to support = {SemCla, PerDis, SemTra, ComMan, CogInt, VisExp, DuaCod, GraEco, CogFit}
Graphical symbol	The visual construct (some v in V) on which the design choice is made.
Visual property	The collection of visual variables of the symbol for which the design choice is made = {Shape, Color, Texture, Size, Brightness, Orientation, Horizontal position (x), Vertical position (y)}
Design Choice	A textual description of the actual design choice made.
Rationale	A textual description of the justification and the evidence supporting it.
Requirement	Explicit requirements from which the design choice followed (if such existed).
Evidence	The level of confidence (and explanation) of the evidence supporting the rationale = {high, moderate, low, very low}

An example is given in Table 6, extracting the design rationale given by Moody [2] for his recommended design for AGENTS in i*. Through such explicit capturing, it is emphasized whether a design choice is justified or not, by making it very clear what evidence it is grounded in.

TABLE 6. EXAMPLE OF A DESIGN CHOICE RATIONALE PROFILE FILLED OUT FOR A SPECIFIC CASE.

Title	Representing agents as James Bond
Principle	Semantic Transparency
Graphical symbol	Agent
Visual property	Shape
Design Choice	Representing an agent as a stick figure wearing sunglasses and holding a pistol: 
Rationale	The stick figure wearing sunglasses and pistol will make people associate the symbol with James Bond, a famous secret ‘agent’.
Requirement	The notation has to be rich to communicate with outsiders.
Evidence	Very low (expert evaluation), based on author’s observation

But what if, like noted in Sec. II.B, we need to capture a rationale for the degree to which we apply a particular theory? Table 7 gives an example for dual coding, relating a decision to ALL symbols of some kind. The rationale presented here lets other people clearly see the degree to which the principle was applied, and whether the design is relevant to their context.

TABLE 7. EXAMPLE OF A DESIGN CHOICE RATIONALE PROFILE FILLED OUT FOR APPLICATION OF DUAL-CODING.

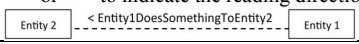

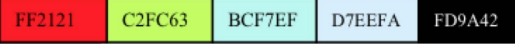

Title	Avoiding writing direction confusion
Principle	Dual coding
Graphical symbol	ALL RELATIONSHIPS
Visual property	Shape, Position
Design Choice	Putting a label on each relationship in the form “subject – verb – object”, encoding it as CamelCase, e.g., “SubjectVerbObject”, and prefix each such label with either “<” or “>” to indicate the reading direction: 
Rationale	By putting a label on each relationship above its graphical representation, people will more quickly understand what they mean. We use these domain models to instruct our programmers to build the systems for us, and the language they use (Pascal, Java) uses CamelCase, adopting the same writing conventions, will make their use simpler and quicker. We explicitly encode reading direction because some of our developers will speak right-to-left languages (e.g., Hebrew, Arabic) during meetings where informal modeling will take place, “on the napkin.” When transferring these models to a final representation, the spatial location and ordering of entities may change, or reverse, leading to potential ambiguities in reading relationships.
Requirement	The notation must be usable in multi-lingual environments.
Evidence	High [14,15] for increased understanding due to redundant coding Moderate (internal survey) for the choice for CamelCase and explicit encoding of reading direction based on an internal survey among our organization’s programmers and their preferred style of coding and documenting.

Table 8 gives an example for important common design choices, here anchoring color palette for the notation, and assigning a pop-out color for a specific element intended to stand out. Note the additional use of evaluation with intended users to ensure appropriateness of the design choice.

TABLE 8. EXAMPLE OF A DESIGN CHOICE RATIONALE PROFILE FILLED OUT TO ENSURE FUNCTIONAL PERCEPTUAL POP-OUT IN A SECURITY LANGUAGE.

Title	Ensuring ‘risk’ elements stand out in security models
Principle	Perceptual discriminability
Graphical symbol	Risk ()
Visual property	Color
Design Choice	Using the following color palette defined in literature [7]:  Using the pop-out color (FF2121) exclusively for ‘risk’ elements, see for example: 
Rationale	For ‘risk’ elements to further stand out besides its unique shape, we want to produce a color pop-out effect. Pop-out effects are achieved when there is sufficient distance in both color and luminance. We will chose a color palette optimized for generating pop-out established in literature, and assign the ‘pop-out’ color value solely to the ‘risk’ elements.
Requirement	The notation must let ‘risk’ elements immediately stand out.
Evidence	Moderate [7] for exact color palette choices Moderate (internal survey) further testing of the palette in survey with intended end-users was favorable, see <data>.

IV. REQUIREMENTS FOR A VERIFIABLE PON APPLICATION: A FRAMEWORK

Based on the analysis in Sec. II, and the requirements derived in Sec. III, we now propose a framework for increasing the verifiability of reporting on PoN-based analyses of notations. The framework is structured so that exact instructions are given what to do for each principle, and where it is important to note requirements for the visual notation that may impact the design choices. For such cases, explicit design rationale according to Sec. II.C is captured to enable readers to understand choices made. For example, if there is a strict requirement that a visual notation is kept as simple as possible, likely there will be more semantic constructs than visual constructs. This would cause symbol overload and negatively impact semiotic clarity.

A. Notation-specific reporting requirements

The following should be reported before any principle-specific analysis is reported:

- List those requirements that impact principle implementation, i.e., to do with types of user, medium of use.
- List all semantic constructs S , visual constructs $V = (\text{Entities } U \text{ Relationships})$, and their mapping, $M: S \rightarrow V$.
- List all used visual variables VisVar and their possible values used in the visual elements V . For example, Shape: square, circle, triangle; Color: red, blue, green.
- List all compositional rules R used for S and V .

Notation
Ensure the following information is reported:
Requirements: <input type="checkbox"/>
S : <input type="checkbox"/> , V : <input type="checkbox"/> , $M: S \rightarrow V$: <input type="checkbox"/> , VisVar : <input type="checkbox"/> , R : <input type="checkbox"/>

B. Principle-specific reporting requirements

For all the principle-specific instructions given below, all results should be explicitly reported. The capture profile from Table 5 should be used for providing design rationale.

Semiotic clarity (SemCla)

Compute the value for symbol ...

... redundancy = $|\{v \in V \mid |\{s \in S: M(s) = v\}| > 1\}|$

... overload $|\{s \in S \mid |\{v \in V: M(s) = v\}| > 1\}|$

... excess = $|\{v \in V \mid \neg \exists s \in S: M(s) = v\}|$

... deficit = $|\{s \in S \mid \neg \exists v \in V: M(s) = v\}|$

For any non-zero value:

Provide a structured design rationale (using the structure provided in Table 5):

Perceptual discriminability (PerDis)

Choose a metric Sim which for every two shapes s, s' of visual constructs $v, v' \in V$ returns their similarity score:

Choose a dissimilarity threshold D if $\text{Sim}(s, s') < D$, meaning that s and s' are not similar:

Provide the following details for each two shapes s, s' of different visual constructs v, v' :

Shape similarity:

Provide $\text{Sim}(s, s') =$

If $\text{Sim}(s, s') \geq D$:

Provide a structured design rationale:

Shape inconsistency

If s is a subtype of s' (e.g., i^* 's POSITION is a subtype of ACTOR) and $\text{Sim}(s, s') < D$:

Provide a structured design rationale:

Discriminability of relationships

If v and v' are relationships and $\text{Sim}(s, s') \geq D$:

Provide a structured design rationale:

Semantic transparency (SemTra)

For each visual construct $v \in V$:

Provide v 's location on the Transparency scale of Fig. 1 and a structured design rationale:

Complexity management (ComMan)

Provide the number of diagram types Types that represent a different level of abstraction:

Is at least one level of abstraction used for *recursive decomposition* of diagrams (i.e., black-boxing, elements (de)compose to new diagrams)?

Yes ☐ No ☐

If the number of levels of abstraction is 0, or *recursive decomposition* is not available:

Provide a structured design rationale:

Cognitive integration (CogInt)
If there are multiple types of diagram types ($ Types > 1$):
For any case in which an s is mapped to different v in different diagrams of types t_1, t_2 ¹ :
Provide a structured design rationale []
Apply cognitive integration of diagrams theory in line with your requirements.
Provide a structured design rationale for the implementation and how it addresses the requirements: []

Visual expressiveness (VisExp)
Provide $ VisVar = []$
Is $ VisVar \geq 8$? Yes [] No []
For the selected visual variables $x \in VisVar$:
Provide a structured design rationale, addressing the extent to which the selected x contribute to: (i) PerDis, (ii) Readability, and (iii) Aesthetics: []

Dual coding (DuaCod)
Apply dual coding theory in line with your requirements.
Provide a structured design rationale for the implementation and how it addresses the requirements: [] (see e.g., Table 8)

Graphic Economy (GraEco)
Provide $ V = []$
If $ Types = 1$ and $ V $ is not 7 ± 2 :
Provide a structured design rationale: []
If $ Types > 1$, for each type t :
$V_t = []$ (where V_t is the set of visual constructs used in diagrams of type t)
Provide $ V_t $ for each $t \in Types$: []
If $ V_t $ is not 7 ± 2 :
Provide a structured design rationale: []

Cognitive Fit (CogFit)
Provide visual dialects in line with your requirements. ²
If relevant in your requirements, provide a dialect for at least different level of expertise and different representational medium.
If no dialects are provided, and your requirements indicate users with different level of expertise, or use in different representational media can be expected, provide a structured design rationale for the lack of dialects: []

V. EXAMPLE APPLICATION OF THE FRAMEWORK

A. The showcase

To demonstrate the *a posteriori* use of the framework, we use it to assess a recent application of the PoN. This can aid in determining points of attention for further improvement of the visual notation. Based on a dataset of PoN applications [12] we selected a recently published domain-specific modeling language for testing environment emulation, TeeVML [22].

TeeVML consists of three levels, each of which has a distinct visual design. While the article [22] is mainly focused on semantics, visual design is noted to be important, and an analysis according to all nine principles of the PoN is given to improve the notation's usability and development productivity. Due to space limitations, we will showcase only one case in this paper, leaving the analysis of a larger corpus of PoN applications to future work. A complete overview of our analysis can be found in an online appendix³.

First, we compare the notation-specific reported data presented in [22] against information required by our framework (see Sec. IV.A), obtaining the following:

TeeVML: Reported data:
Requirements: <input type="checkbox"/>
S: <input checked="" type="checkbox"/> , V: <input checked="" type="checkbox"/> , M: S \rightarrow V: <input checked="" type="checkbox"/> , VisVar: <input type="checkbox"/> , R: <input type="checkbox"/>

While only visual variables are explicitly listed, a listing of compositional rules is not provided. Yet the presented overviews of S, V and their mapping enables readers to verify several principles.

Due to space limitations, we focus here on two principles. The first is graphic economy, which is noted in [22] to be a key consideration. While $|V|$ values were not given, they can be calculated based on S, and shown to be within the 7 ± 2 range for almost all diagram types. However, since R is missing, we cannot infer whether diagram types can be linked, and thus shown at the same time. It is therefore not possible to determine whether the overall $|V|$ of 26 is a major consideration.

TeeVML: Graphic Economy (GraEco)
$ V = [26]$
As $ Types = 3$, we consider V_t for each $t \in Types$:
$t_1 = \text{Signature}, V_{t_1} = [9]$
$t_2 = \text{Protocol}, V_{t_2} = [7]$
$t_3 = \text{Behavior}, V_{t_3} = [10]$
Provided design rationale: ["A key design consideration is to minimize the number of visual symbols." [22]]

The second principle we focus on is visual expressiveness. It is noted that "shape, color, orientation, texture, etc." were used. This raises two concerns: (i) it is not clear exactly how many, and which variables were used in total, and (ii) it is not clear how the combination of variables is supposed to contribute to PerDis, readability, and aesthetics. Thus we obtain:

¹ This rationale should be in line with the not satisfying semiotic clarity.

² New design choices in dialects should be re-evaluated with the framework.

³ See www.dirkvanderlinden.eu/data/ for the complete assessment.

TeeVML: Visual expressiveness (VisExp)
VisVar = [at least 4]
Provided rationale for contribution of used $x \in \text{VisVar}$ to PerDis, Readability, Aesthetics:
Shape: [?] Color: [?] Orientation: [?] Texture: [?]

It should be noted the above fragment of our framework application highlights gaps in reporting the PoN application to a visual design, and not weaknesses of the latter. As the PoN does not provide enough guidance on *how* to report on this rationale, the aim of our framework is to address this gap by pointing out exactly these points where further elaboration and design rationale are needed. By the framework’s stimulation of capturing explicit rationale grounded in evidence, designers using it would be guided to spend more time understanding the impact of their design choices on their visual notation’s users as well as provide more comprehensive and convincing reports regarding the design choices they make.

B. Deriving required points of attention

From the full analysis of the above showcase, a number of points to improve *the verifiability* of the design choices (as opposed to the choices themselves) made by notation designers can be considered. To illustrate how a *posteriori* application of the framework can be useful in strengthening existing notation’s design, some of these points are summarized below:

- Provide (a link to) an exhaustive listing of the rules of composition for S and V. This would also aid in the verifiability of certain principles, such as complexity management, by establishing exactly how the information hiding mechanism works. Cognitive integration’s verifiability also requires these composition rules.
- Provide a description of requirements that impacted the principles implementation, in order to assess, e.g., the relative importance of your dual coding implementation, whether cognitive integration is sufficient for your users.
- *Principle-specific:*
 - Provide explicit scores for symbol redundancy, overload, excess, and deficit.
 - Provide a similarity metric *Sim* used to verifiably assess that shapes are distinct.
 - Provide design rationale for the design of each icon meant to be semantically transparent.
 - Provide |VisVar| and assess each x in |VisVar|’s impact on PerDis, readability, and aesthetics.
 - Provide a design rationale for the used style of dual coding, in particular on the complementary use of redundant label coding and textual instantiation encoding.
 - Elaborate from the requirements whether the three diagrams are used together, and if so, provide a design rationale for the overall |V| of 26.

Some of the noted points of attention could likely be intended design to satisfy a requirement for the notation. If this is the case, explicitly capturing that reason would make it verifiable that the principle was not left unsatisfied.

VI. DISCUSSION & CONTRIBUTION

The use of an evidence-based approach with focus on design rationale might seem counter-intuitive for a design theory. The PoN is at first a normative Type V design theory [23] – that is, it is meant to design new artifacts to solve existing problems. Typically, argumentation has been used as a core factor of validation in such research efforts. The presentation of a strong argument why an artifact (e.g., a method, tool, model) helps with something is its core validation. However, if we take into account that many applications of the PoN lack detail in their reporting [12], it seems that the presentation of arguments as to why a visual notation satisfies a principle is not simple for many designers. The worth of design research is established by its validity to users, novelty of the artifact, and importantly: the *persuasiveness* of the given evidence [24]. With the lack of detail in reporting of many PoN applications, there is little persuasiveness that their claims to being cognitively effective truly hold – which is exactly where the evidence-driven focus aids designers by being forced to think about, and accurately present this data. Whether such evidence is typically of lower level of confidence according to our model is not a problem directly, as we want to first simply compare evidence for different designs, and also stimulate designers to further improve their validity.

The utility of our framework should thus be considered for two distinct audiences: (i) designers of visual notations, and (ii) (potential) users of visual notations. For the former, we stimulate better design and reporting as a core utility, and for the latter, we provide utility by enabling them to critically assess the quality and proven usefulness of a visual notation they might want to use for a specific task. As seen from the showcase in Section V, some visual notations designed with the PoN in mind cannot simply be said to truly satisfy its normative instructions. Thus, it is not trivial to say that a visual notation claimed to have been designed according to the PoN is indeed cognitively effective, or well designed at all. From our work of surveying PoN applications [12], we found that this scenario is fairly typical. The proposed framework might thus help to raise the level of verified design by making it more visible to users of notations how well designed they can truly claim to be.

We designed the framework in such a way that it translates to a checklist-style approach, so that designers of a visual notation are not needlessly encumbered by having to write additional formalizations. It can thus also be easily integrated into approaches for systematically ordering when to apply which principle, such as PoN-S [3]. The main complexity of use manifests in the need to be explicit and convincing about the rationales used in order to justify design choices – and the need to present and classify the evidence supporting them. We further envision the framework’s automation, into a *computer-aided design* tool, helping designers by providing relevant feedback on what principles are left unaddressed or unsatisfied, and guiding them in providing all the information needed.

As hinted at from the showcase in Sec. V.B., we see a use for a *posteriori* analysis – applying it to the full corpus of PoN applications. Doing so would lead to a general overview of what aspects are well- and under-reported, what principles are backed up by the most evidence, as well as what kinds of evi-

dence is used in general. Furthermore, establishing such a dataset would reinforce work based on re-using validated design elements [25] by allowing to identify and reuse those elements with the highest level of evidence already in existence.

VII. CONCLUDING OUTLOOK

This article presented a systematic framework to be used to guide the design of visual notations according to the Physics of Notations theory, as well as a *a posteriori* analysis of PoN-guided designs. The novelty of our framework is that it explicitly acknowledges the inherent limits of operationalization for different principles (requiring user involvement, interpretation of theory, etc.), and focuses on guiding designers to make their design choices explicit and grounded in evidence and in requirements for their notation. Our framework proposes a systematic way to capture evidence-based design rationale, and incorporates a set of guidelines for designers to encode all information needed for others to understand and contextualize the design choices made.

We intend to implement the framework as a computer-aided assessment tool, and further apply it to generate an exhaustive dataset of PoN applications, showcasing what principles are typically most verified and where evidence typically lacks, to further guide designers as to which principles require additional consideration to ensure their full application.

REFERENCES

- [1] A. D. Oxman and G. W. Group, "Grading quality of evidence and strength of recommendations," *Bmj*, vol. 328, pp. 1490-4, 2004.
- [2] D. L. Moody, P. Heymans, and R. Matulevičius, "Visual syntax does matter: improving the cognitive effectiveness of the i* visual notation," *Requirements Engineering*, vol. 15, pp. 141-175, 2010.
- [3] G. A. Miller, "The magical number seven, plus or minus two: some limits on our capacity for processing information," *Psychological review*, vol. 63, p. 81, 1956.
- [4] D. L. Moody, "The 'Physics' of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering," *Ieee Transactions on Software Engineering*, vol. 35, pp. 756-779, Nov-Dec 2009.
- [5] M. d. G. da Silva Teixeira, G. K. Quirino, F. Gailly, R. de Almeida Falbo, G. Guizzardi, and M. P. Barcellos, "PoN-S: A Systematic Approach for Applying the Physics of Notation (PoN)," in *International Workshop on Business Process Modeling, Development and Support*, 2016, pp. 432-447.
- [6] H. Störrle and A. Fish, "Towards an Operationalization of the 'Physics of Notations' for the Analysis of Visual Languages," in *Model-Driven Engineering Languages and Systems*, ed: Springer, 2013, pp. 104-120.
- [7] J. Stark, R. Braun, and W. Esswein, "Systemizing Colour for Conceptual Modeling," presented at the 13th International Conference on Wirtschaftsinformatik, St. Gallen, Switzerland, 2017.
- [8] D. Granada, J. M. Vara, V. A. Bollati, and E. Marcos, "Enabling the Development of Cognitive Effective Visual DSLs," in *International Conference on Model Driven Engineering Languages and Systems*, 2014, pp. 535-551.
- [9] D. van der Linden, A. Zamansky, and I. Hadar, "How cognitively effective is a visual notation? on the inherent difficulty of operationalizing the physics of notations," in *International Workshop on Business Process Modeling, Development and Support*, 2016, pp. 448-462.
- [10] D. van der Linden and I. Hadar, "User involvement in applications of the PoN," in *International Conference on Advanced Information Systems Engineering*, 2016, pp. 109-115.
- [11] T.-F. Kummer, J. Recker, and J. Mendling, "Enhancing understandability of process models through cultural-dependent color adjustments," *Decision Support Systems*, vol. 87, pp. 1-12, 2016.
- [12] D. van der Linden and I. Hadar, "Evaluating the Evaluators - An Analysis of Cognitive Effectiveness Improvement Efforts for Visual Notations," presented at the Proceedings of the 11th International Conference on Evaluation of Novel Software Approaches to Software Engineering, Rome, Italy, 2016.
- [13] J. Kim, J. Hahn, and H. Hahn, "How do we understand a system with (so) many diagrams? Cognitive integration processes in diagrammatic reasoning," *Information Systems Research*, vol. 11, pp. 284-303, 2000.
- [14] A. Paivio, *Imagery and verbal processes*: Psychology Press, 2013.
- [15] A. Paivio, *Mental representations: A dual coding approach*: Oxford University Press, 1990.
- [16] I. Vessey, "Cognitive fit: A theory - based analysis of the graphs versus tables literature," *Decision Sciences*, vol. 22, pp. 219-240, 1991.
- [17] J. Lee and K.-Y. Lai, "What's in design rationale?," *Human-Computer Interaction*, vol. 6, pp. 251-280, 1991.
- [18] F. M. Shipman and R. J. McCall, "Integrating different perspectives on design rationale: Supporting the emergence of design rationale from design communication," *Artificial Intelligence for Engineering, Design, Analysis and Manufacturing*, vol. 11, pp. 141-154, 1997.
- [19] D. L. Sackett, "Rules of evidence and clinical recommendations on the use of antithrombotic agents," *CHEST Journal*, vol. 95, pp. 2S-4S, 1989.
- [20] G. H. Guyatt, A. D. Oxman, G. E. Vist, R. Kunz, Y. Falck-Ytter, P. Alonso-Coello, *et al.*, "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations," *BMJ: British Medical Journal*, vol. 336, p. 924, 2008.
- [21] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and software technology*, vol. 50, pp. 833-859, 2008.
- [22] J. Liu, J. Grundy, I. Avazpour, and M. Abdelrazek, "A Domain-Specific Visual Modeling Language for Testing Environment Emulation," in *2016 IEEE Symposium on Visual Languages and Human-Centric Computing*, Cambridge, UK, 2016.
- [23] S. Gregor, "The nature of theory in information systems," *MIS quarterly*, pp. 611-642, 2006.
- [24] R. H. Von Alan, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS quarterly*, vol. 28, pp. 75-105, 2004.
- [25] D. van der Linden, I. Hadar, and A. Zamansky, "Towards a Marketplace of Visual Elements for Notation Design," in *Requirements Engineering Conference (RE), 2016 IEEE 24th International*, 2016, pp. 353-358.